

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2002, 16:22:02 ; Search time 1986.52 Seconds

(without alignments)
3286.692 Million cell updates/sec

Title: US-09-880-887-9

Perfect score: 312

Sequence: 1 gttgtttatgcatccttt.....cgtatcttttaccattcag 312

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_to:*
10: gb_sy:*
11: gb_sy:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Score Length DB ID Description

1	312	100.0	312	6	AR163685	AR163685 Sequence
2	312	100.0	312	6	BD000102	BD000102 Modified
3	146.2	46.9	38059	9	HDMFTXC	K02402 Human coagu
4	146.2	46.9	153615	2	AL645665	AL645665 Homo sapi
5	146.2	46.9	158557	9	HS88D7	AL033403 Human sapi
6	143	45.8	822	9	AB062458S2	AB062460 Pan trogl
7	142.8	45.8	599	9	HDMFTXC1	K02049 Human facto
8	142.2	45.6	600	9	HDMFTXC1	K02049 Human facto
9	142	45.5	822	9	AB062459S2	AB062461 Pan trogl
10	117.2	37.6	347	9	AB062458S1	AB062458 Pan trogl
11	117.2	37.6	347	9	AB062459S1	AB062459 Pan trogl
12	115	36.9	10329	6	AX347024	AX347024 Sequence
13	103.6	33.2	10329	6	AX347025	AX347025 Sequence
14	67.4	21.6	221406	2	AL671984	AL671984 Mus muscu
15	51.8	16.6	9927	6	AX345014	AX345014 Sequence
16	47.6	15.3	77585	9	AL133418	AL133418 Human DNA
17	47.6	15.3	154935	2	AL356781	AL356781 Homo sapi
18	47.6	15.3	168584	2	AL391994	AL391994 Homo sapi
19	47.6	15.3	251039	2	AL592290	AL592290 Homo sapi
20	42.6	13.7	15980	3	CCA242872	AJ742872 Ceratilis
21	42.2	13.5	38980	3	CER06C1	Z81106 Caenorhabd
22	42	13.5	42223	3	CER46G10	Z50177 Caenorhabd
23	42	13.5	150879	2	AC074017	AC074017 Homo sapi
24	42	13.5	165657	30	AC074341	AC074341 Homo sapi
25	42	13.5	166657	2	AC104161	AC104161 Homo sapi
26	42	13.5	173794	9	AC023108	AC023108 Homo sapi
27	41.6	13.3	2435	1	ACCPQ01T1	D500330 Acinetobact
28	41.6	13.3	50946	8	AC007212	AC007212 Arabidops
29	41.4	13.3	136006	9	AC006151	AC006151 Homo sapi
30	41.2	13.2	120336	2	AC108697	AC108697 Homo sapi
31	41.2	13.2	148387	2	AC107015	AC107015 Homo sapi
32	41.2	13.2	153336	2	AC025020	AC025020 Homo sapi
33	41.2	13.2	163022	2	AC016851	AC016851 Homo sapi
34	41	13.1	166689	2	AC058815	AC058815 Homo sapi
35	41	13.1	198217	9	AL354942	AL354942 Human DNA
36	40.8	13.1	36286	3	CEY105EBC	AF099923 Caenorhab
37	40.6	13.0	91819	3	CEY105EBC	AF099923 Caenorhab
38	40.6	13.0	110000	3	CEY105EBC_4	AF099923 Caenorhab
39	40.6	13.0	277607	2	CEY105EBC_4	Continuation (5 of
40	40.2	12.9	186274	2	AC010341	AL132876 Caenorhab
41	40	12.8	143171	8	CLECGA	AC010341 Homo sapi
42	39.8	12.8	38066	3	U97549	X70810 Euglena gra
43	39.8	12.8	151821	2	AC098213	U97549 Caenorhabd
44	39.8	12.8	158502	9	AC005150	AC098213 Rattus no
45	39.4	12.6	1760	6	E12779	AC005150 Homo sapi

ALIGNMENTS

.

RESULT 1
AR163685
LOCUS AR163685 312 bp DNA
SEQUENCE AR163685 9 from patent US 6271025.
ACCESSION AR163685
VERSION AR163685.1 GI:16234366
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 312)
AUTHORS Negrier C. and Plantier J. Luc.
TITLE Modified factor VIII cDNA
JOURNAL Patent: US 6271025-A 9 07-ANG-2001;
FEATURES Location/Qualifiers
source 1..312
BASE COUNT 96 a 47 c 53 g 116 t
ORIGIN
Query Match 100.0%; Score 312; DB 6; Length 312;
Best Local Similarity 100.0%; Pred. No. 9,4e-58;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgttatcatcctcttttaataacatcatgagtagtgccttgctttagatataaaga 60
 Db 1 GTTGTATTATGCATCCCTTTTAAATACATGAGTATGCTTGCTTTTGAATATGAGAA 60

QY 61 tatctgatgctgtcttcttcaataatttgattacatgatttgacagcaatatgaaga 120
 Db 61 TATCTGATGCTGTCTTCTTCACTAAATTTGATTTACATGATTGACAGCAATATTGAGA 120

QY 121 gtctaaagccagcagcagaggttgtagtactgtgggaacatcacatcagatttggcccca 180
 Db 121 GTCTAACGCCAGCAGCAGAGGTGTGTAAGTACTGTGGAACTCACAGATTTTGGCTCCA 180

QY 181 tgccttaagagaataatgcttcacagattatgatttaaaacaagaacttcttaaga 240
 Db 181 TGCCCTTAAGAGAAATTTGCTTTCAGATTTATTGATTTAAACAAAGACTTTCTTAGA 240

QY 241 gatgtaaaatttcatgagtcttcttcttcttcttcttcttcttcttcttcttcttct 300
 Db 241 GATGTAAATTTTCATGATGATGTTCTTTTTCCTAAACTAAAGAAATTAACGCGTATTC 300

QY 301 tttaatttcag 312
 Db 301 TTACATTTCAG 312

RESULT 2
 BD000102 312 bp DNA linear PAT 31-JAN-2002

LOCUS BD000102 Modified factor VIII cDNA.
 DEFINITION BD000102
 ACCESSION BD000102.1 GI:18623181
 VERSION JP 2000287694-A/9.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 312)
 AUTHORS Negrie C. and Puntantje J.R.
 TITLE Modified factor VIII cDNA
 JOURNAL Patent: JP 2000287694-A 9 17-OCT-2000;
 COMMENT ABEENTIS BERINGU GMBH
 OS Homo sapiens (human)
 PN JP 2000287694-A/9
 PD 17-OCT-2000
 PE 16-MAR-2000 JP 2000073665
 PR 17-MAR-1999 EP 99104050:2
 PI CLAUDE NEGRIE, JEAN RYUOYU PUNANTJE
 PC C12N15/09, A61K48/00, A61P7/04, C12P21/02, C12N5/10, C12P21/02,
 PC C12N1:91',
 PC C12N15/00, C12N5/00
 CC
 FH
 FT source Key Location/Qualifiers
 1. 312 /organism='Homo sapiens (human)'.
 FEATURES
 source 1. 312 Location/Qualifiers
 BASE COUNT 96 a 47 c 53 g 116 t
 ORIGIN

Query Match 100.0%; Score 312; DB 6; Length 312;
 Best Local Similarity 100.0%; Pred. No. 9, 4e-58;
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgttatcatcctcttttaataacatcatgagtagtgccttgctttagatataaaga 60
 Db 1 GTTGTATTATGCATCCCTTTTAAATACATGAGTATGCTTGCTTTTGAATATGAGAA 60

QY 61 tatctgatgctgtcttcttcaataatttgattacatgatttgacagcaatatgaaga 120

Db 61 TATCTGATGCTGTCTTCTTCACTAAATTTGATTTACATGATTGACAGCAATATTGAGA 120

QY 121 gtctaaagccagcagcagaggttgtagtactgtgggaacatcacatcagatttggcccca 180
 Db 121 GTCTAACGCCAGCAGCAGAGGTGTGTAAGTACTGTGGAACTCACAGATTTTGGCTCCA 180

QY 181 tgccttaagagaataatgcttcacagattatgatttaaaacaagaacttcttaaga 240
 Db 181 TGCCCTTAAGAGAAATTTGCTTTCAGATTTATTGATTTAAACAAAGACTTTCTTAGA 240

QY 241 gatgtaaaatttcatgagtcttcttcttcttcttcttcttcttcttcttcttcttct 300
 Db 241 GATGTAAATTTTCATGATGATGTTCTTTTTCCTAAACTAAAGAAATTAACGCGTATTC 300

QY 301 tttaatttcag 312
 Db 301 TTACATTTCAG 312

RESULT 3
 HUMFIX 38059 bp DNA linear PRI 30-APR-1996

LOCUS HUMFIX Human coagulation factor IX gene, complete cds.
 DEFINITION K02402
 ACCESSION K02402.1 GI:182612
 VERSION K02402.1
 KEYWORDS Alu repeat; Christmas factor; Kpni repetitive sequence; antihemophilic factor B; factor IX; repeat region; simple repetitive sequence.
 SOURCE Homo sapiens (clone: FIX-lambda-[6,36,53,61].) (tissue library: T. Maniatis et al.) DNA; and Homo sapiens (clone: FIX-lambda-4243) DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 38059)
 AUTHORS Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W. and Kurachi, K.
 TITLE Nucleotide sequence of the gene for human factor IX (antihemophilic factor B)
 JOURNAL Biochemistry 24 (14), 3736-3750 (1985)
 MEDLINE 86000558
 REFERENCE 2 (bases 23487 to 23556)
 AUTHORS Rees, D.J., Rizza, C.R. and Brownlee, G.G.
 TITLE Haemophilia B caused by a point mutation in a donor splice junction of the human factor IX gene
 JOURNAL Nature 316 (6029), 643-645 (1985)
 MEDLINE 85296286
 REFERENCE 3 (bases 23378 to 23387)
 AUTHORS Graham, J.B., Lubahn, D.B., Lord, S.T., Kirshtein, J., Nilsson, I.M., Wallmark, A., Ljung, R., Frazier, L.D., Ware, J.L., Lin, S.W., Stafford, D.W. and Bosco, J.
 TITLE The Malmo polymorphism of coagulation factor IX, an immunologic polymorphism due to dimorphism of residue 148 that is in linkage disequilibrium with two other F.IX polymorphisms
 JOURNAL Am. J. Hum. Genet. 42 (4), 573-580 (1988)
 MEDLINE 88161064
 REFERENCE 4 (sites)
 AUTHORS Hirosewa, S., Fahner, J.B., Salier, J.-P., Wu, C.-T., Lovrien, E. and Kurachi, K.
 TITLE Structural and functional basis of the developmental regulation of human factor IX gene: factor IX Leyden
 JOURNAL Unpublished (1990)
 COMMENT Sequence for [1] kindly submitted on floppy by K. Kurachi, 05-AUG-1985.
 [1] notes a potential TATA box (2939-2942) and polyadenylation signal (35701-35706); and notes two start codons (downstream of the start codon annotated below) that may be alternative and/or preferred starts for the factor IX prepropeptide. Several tracts of simple repetitive sequence are present [1], including regions with the potential for hairpin and/or Z-DNA formation. [1] describes six long open reading frames in the intron and on the complementary strand.
 Location/Qualifiers

```
source
1. .38059
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="xq26.3-q27.1"
/clone="FIX-lambda-4243"
/clone="FIX-lambda-[6,36,53,61]."
/cell_line="49,xxxxx"
/cell_type="fibroblast"
/tissue_idb="T.Manialis et al."
1. .2596
/rep_family="KpnI"
prim_transcript
2966. .35722
/gene="F9"
/gene="F9 mRNA"
2966. .35722
/gene="F9"
2966. .3082
/gene="F9"
/number=1
/note="G00-119-900"
CDS
join(2995. .3082,9291. .9454,9643. .9667,13357. .13470,
20634. .20762,23328. .23530,33004. .33118,33787. .34334)
/gene="F9"
/note="precursor"
/codon_start=1
/product="factor IX"
/protein_id="AAB59620.1"
/db_xref="GI:182613"
/db_xref="GDB:G00-119-900"
/translation="MQRVNMIMASPOLITICILGYLLSAECTVFLDHNANKILNP
KRYSGKLEEVGNLEFECMEKESPEEARVEVENTERTTEFMKQYVDGDOCESNG
LNGSGKDDINRYCWCPCPGEGKNCILDVTCNKGRCGFCNKSADNKVVCSCNG
YRLAENKSCERAPFPCGRVSVSQTSLTRAPEVDVYVNSTEATLLDITOST
QSENDTRVVGEDAPGQFPQVYVNLGKYDAFCGSIWNEKMTVTAHCEVGVKIT
VAGEHNIETEHTQKRNVIIRIIPHNYAAINKYDIALLDELPIVLNSYVPI
CIADKETNIFLKGSGYVSGWGFPHKGRSALVQLRYPLVDRATCLRSKFTIYN
NMFCAFHGEGRDCQDSCGPHVTEVGTSPILGLISWCECAMKGKYLIVKVSX
VNWIKERTKLT"
join(2995. .3082,9291. .9340)
sig_peptide
/gene="F9"
/note="G00-119-900"
3083. .9290
/gene="F9"
/note="G00-119-900"
/number=1
7298. .7593
/rep_family="Alu"
8469. .8520
/gene="F9"
/note="t. .50 bp. .a in FIX-lambda-4243; ta in
FIX-lambda-[36,61]; G00-119-900"
/citation=[1]
/replace="ta"
9291. .9454
/gene="F9"
/note="G00-119-900"
/mat_peptide
join(9341. .9454,9643. .9667,13357. .13470,20634. .20762,
23328. .23380)
/gene="F9"
/note="G00-119-900"
/product="factor IX light chain"
9455. .9642
/gene="F9"
/number=2
/note="G00-119-900"
9643. .9667
/gene="F9"
/note="G00-119-900"
/number=3
9668. .13356
/gene="F9"
/note="G00-119-900"
intron
variation
/number=3
10041
/gene="F9"
/note="g in one allele; c in another allele (loss of XmnI
recognition pattern); G00-119-900"
/replace="c"
13357. .13470
/gene="F9"
/note="G00-119-900"
/number=4
13471. .20633
/gene="F9"
/note="G00-119-900"
/number=4
14076. .14079
/gene="F9"
/note="tcga in one allele; nnn in another allele (loss of
RagI site); G00-119-900"
/replace="nnn"
18165. .20265
/rep_family="KpnI"
20416
/gene="F9"
/citation=[4]
/replace="a"
20634. .20762
/gene="F9"
/note="G00-119-900"
/number=5
20763. .23327
/gene="F9"
/note="G00-119-900"
/number=5
23328. .23530
/gene="F9"
/note="G00-119-900"
/number=6
23387
/gene="F9"
/note="g in [1]; a in [3] Ala->Thr; G00-119-900"
/replace="a"
23387
/gene="F9"
/note="g in one allele; a in another allele"
/replace="a"
join(23486. .23530,33004. .33118,33787. .34331)
/gene="F9"
/note="G00-119-900"
/product="factor IX heavy chain"
23531. .33003
/gene="F9"
/note="G00-119-900"
/number=6
23531
/gene="F9"
/note="g in [1]; t in [2] (haemophilia patient);
G00-119-900"
/replace="t"
24172. .24475
/rep_family="Alu"
25863. .26091
/rep_family="Alu"
31537. .31809
/rep_family="Alu"
33004. .33118
/gene="F9"
/note="G00-119-900"
/number=7
33119. .33786
/gene="F9"
/note="G00-119-900"
/number=7
33787. .35722
exon
variation
/number=3
10041
/gene="F9"
/note="g in one allele; c in another allele (loss of XmnI
recognition pattern); G00-119-900"
/replace="c"
13357. .13470
/gene="F9"
/note="G00-119-900"
/number=4
13471. .20633
/gene="F9"
/note="G00-119-900"
/number=4
14076. .14079
/gene="F9"
/note="tcga in one allele; nnn in another allele (loss of
RagI site); G00-119-900"
/replace="nnn"
18165. .20265
/rep_family="KpnI"
20416
/gene="F9"
/citation=[4]
/replace="a"
20634. .20762
/gene="F9"
/note="G00-119-900"
/number=5
20763. .23327
/gene="F9"
/note="G00-119-900"
/number=5
23328. .23530
/gene="F9"
/note="G00-119-900"
/number=6
23387
/gene="F9"
/note="g in [1]; a in [3] Ala->Thr; G00-119-900"
/replace="a"
23387
/gene="F9"
/note="g in one allele; a in another allele"
/replace="a"
join(23486. .23530,33004. .33118,33787. .34331)
/gene="F9"
/note="G00-119-900"
/product="factor IX heavy chain"
23531. .33003
/gene="F9"
/note="G00-119-900"
/number=6
23531
/gene="F9"
/note="g in [1]; t in [2] (haemophilia patient);
G00-119-900"
/replace="t"
24172. .24475
/rep_family="Alu"
25863. .26091
/rep_family="Alu"
31537. .31809
/rep_family="Alu"
33004. .33118
/gene="F9"
/note="G00-119-900"
/number=7
33119. .33786
/gene="F9"
/note="G00-119-900"
/number=7
33787. .35722
exon
```

```
/gene="F9"
/notes="G00-119-900"
/number=8

Query Match      46.9%; Score 146.2; DB 9; Length 38059;
Best Local Similarity 81.6%; Pred. No. 3.3e-22;
Matches 169; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Oy 103 tgcagcaatctgaagagctaacagccagcagcaggtgtgtaagctctgtggaaca 162
Db 9088 TAAAGTAAATTTGAATTTTAAATTCCTCAATTCATGCTATACAGTACTGTGGGAACA 9147

Oy 163 tcacagatttggtccatgccctaaagagaatltgcttcagattatgtgataaa 222
Db 9148 TCACAGATTTTGGCTCATGCCCTAAAGAGAAATTTGGCTTCAGATTTATTTGGATTAAAA 9207

Oy 223 acaaaagcttcttaagagatgtcaaaatttcattgattgtttcttttgcataaactaa 282
Db 9208 ACAAAAGCTTTCTTAAAGATGTAATAATTTTCATGATGTTCTTTTGTCTAAACTAA 9267

Oy 283 agaattacgcgtattctttacattt 309
Db 9268 AGAATTATTTCTTTTACATTTCAGTTT 9294
```

```
RESULT 4
AL645665 153615 bp DNA linear HTG 01-FEB-2002
LOCUS Homo sapiens chromosome X clone RP11-963P9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL645665
VERSION AL645665.9 GI:18491387
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
Heath, P.
Direct Submission
Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Feb 4, 2002 this sequence version replaced gi:18476880.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba963p9
----- Summary Statistics
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 152173 bases at least Q40
Consensus quality: 152586 bases at least Q30
Consensus quality: 152764 bases at least Q20
Insert size: 153415; sum-of-ctgts
Insert size: 144208; 38.0% error; agarose-fp
Quality coverage: 20.53x in Q20 bases; sum-of-ctgts Quality
coverage: 22.15x in Q20 bases; agarose-fp
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 28734: contig of 28734 bp in length
* 28733 28834: gap of 100 bp
```

```
* 28835 44671: contig of 15837 bp in length
* 44672 44771: gap of 100 bp
* 44772 153615: contig of 108844 bp in length.

FEATURES
Source
1..153615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-963P9"
/clone_1bp="RPCR-11.4"
1..28734
/notes="assembly_fragment:06384
fragment_chain:1
clone_end:SP6
vector_side:left"
28835..44671
/notes="assembly_fragment:02107
fragment_chain:1"
44772..153615
/notes="assembly_fragment:00800
fragment_chain:1"

BASE COUNT 50474 a 29730 c 28392 g 44819 t 200 others
ORIGIN
```

```
Query Match      46.9%; Score 146.2; DB 2; Length 153615;
Best Local Similarity 81.6%; Pred. No. 2.8e-22;
Matches 169; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Oy 103 tgcagcaatctgaagagctaacagccagcagcaggtgtgtaagctctgtggaaca 162
Db 85574 TAAAGTAAATTTGAATTTTAAATTCCTCAATTCATGCTATACAGTACTGTGGGAACA 85633

Oy 163 tcacagatttggtccatgccctaaagagaatltgcttcagattatgtgataaa 222
Db 85634 TCACAGATTTTGGCTCATGCCCTAAAGAGAAATTTGGCTTCAGATTTATTTGGATTAAAA 85693

Oy 223 acaaaagcttcttaagagatgtcaaaatttcattgattgtttcttttgcataaactaa 282
Db 85694 ACAAAAGCTTTCTTAAAGATGTAATAATTTTCATGATGTTCTTTTGTCTAAACTAA 85753

Oy 283 agaattacgcgtattctttacattt 309
Db 85754 AGAATTATTTCTTTTACATTTCAGTTT 85780
```

```
RESULT 5
HS88D7 158557 bp DNA linear PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 88D7 on chromosome Xq25-26.3 Contains
DEFINITION F9 (coagulation factor IX (plasma thromboplastic component,
Christmas disease, haemophilia B)), db1 oncogene, EST, STS, GSS,
complete sequence.
ACCESSION AL033403
VERSION AL033403.1 GI:3859054
KEYWORDS HTG; christmas factor; db1 oncogene; F9; factor IX; glycoprotein;
SOURCE oncogene; phosphoprotein; proto-oncogene.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 158557)
REFERENCE 1
AUTHORS Bird, C.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
```

This sequence is the entire insert of clone 88D7.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/Heb/Chrx>
88D7 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://dacpac.med.buffalo.edu/VECTOR.pDACA>.

FEATURES

```
source
1..158557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q25-26.3"
/clone.lib="RPC1-6"
2..3009
/feature="L1PA2 repeat: matches 3136. .6144 of consensus"
3023..3684
/feature="L1M4 repeat: matches 4639. .5306 of consensus"
3693..3775
/feature="L1MC/D repeat: matches 5517. .5601 of consensus"
join(<4083..4171,10329..10492,10681..10705,14395..14508,
21682..21810,24381..24583,34060..34174,34843..36777)
/feature="P9"
/feature="match: DNA M21757 J00137 A13997 M3826 M11309
J00136 A22493 M23109 M35672 U51135 M26247 M26233 M26234
M26235 M26236 M26237 Z50155; match: SRS G28615"
/feature="not_experimental"
4083..36777
/feature="P9"
join(<4099..4171,10329..10492,10681..10705,14395..14508,
21682..21810,24381..24583,34060..34174,34843..35390)
/feature="P9"
/feature="match: protein P00740 P19540 P16294 P00741 P16293
P16296 P16295 P16292 P16291 P25155"
/codon_start=1
/feature="not_experimental"
/feature="P9 (coagulation factor IX (plasma thromboplastic
component, Christmas disease, haemophilia B))"
/protein_id="CAA21954.1"
/db_xref="GI:3859055"
/db_xref="SPTREMBL:Q14316"
/translation="MIMASPLITICLGYLSAECTVPLDHNANKILNPKRYS
GKLEFFVGNLEREMERKCSFEAREFEVNTERTTERPKQVYDGDCESPCLNGS
CKDINSTECHCPREFEGKNCLELDYTCIKNGKREDFCKNSADNKKVSCSEGRLE
NOKSCGPAVPPCGSVYSOTSKLATRATVPDVTVAETILNDITOSTSFND
FTRVVGSDAKPGQFPMQVYNGKVDACGSIYNEKVIYAAACVEGVIYVAGE
HNIEETHEEOKRNVYRIIPHHNYAAINKYNIHDALELDEPLVNSYPTICAK
EYNIETLEKGGVYSGMGVFRGSAIVLYLVPLVDATCTCRSKRTYNNMFAK
GEHGGRCDSGCGRHVTEVEGTSPLTIGIISGECAMKGIYTKVSRVYMWIK
EKTKLT"
5799..6479
/feature="L2 repeat: matches 1963. .2708 of consensus"
6562..6807
/feature="MIR repeat: matches 6. .262 of consensus"
6829..6925
/feature="L2 repeat: matches 1641. .1739 of consensus"
6965..7015
/feature="L2 repeat: matches 2652. .2700 of consensus"
7150..7512
/feature="L2 repeat: matches 1244. .1617 of consensus"
7565..7764
/feature="MIR repeat: matches 14. .198 of consensus"
7771..8111
/feature="MER2 repeat: matches 1. .345 of consensus"
8388..8685
```

```
repeat_region
9091..9431
/feature="MER7A repeat: matches 2. .346 of consensus"
9468..9609
/feature="71 copies 2 mer at 78% conserved"
10959..11604
/feature="L2 repeat: matches 2123. .2746 of consensus"
12270..12305
/feature="L18 copies 2 mer aa 81% conserved"
12516..12693
/feature="MIR repeat: matches 9. .190 of consensus"
14736..14864
/feature="MIR repeat: matches 52. .193 of consensus"
17136..17189
/feature="L2 repeat: matches 2645. .2706 of consensus"
17940..18249
/feature="L2R16C repeat: matches 58. .383 of consensus"
18718..19054
/feature="L1MEC repeat: matches 2414. .2414 of consensus"
19251..19679
/feature="L1PA14 repeat: matches 5718. .6149 of consensus"
19857..20579
/feature="L1M4 repeat: matches 4181. .4940 of consensus"
20625..20789
/feature="L1M4 repeat: matches 5079. .5249 of consensus"
20788..21386
/feature="L1ME3 repeat: matches 5517. .6155 of consensus"
22684..23154
/feature="MER9 repeat: matches 1. .511 of consensus"
25061..25226
/feature="MER5A repeat: matches 16. .186 of consensus"
25227..25532
/feature="AluX repeat: matches 1. .302 of consensus"
25840..26048
/feature="MER3 repeat: matches 2. .208 of consensus"
26919..27152
/feature="AluY repeat: matches 69. .302 of consensus"
31078..32008
/feature="MER45C repeat: matches 1. .948 of consensus"
32082..32243
/feature="MER5A repeat: matches 10. .182 of consensus"
32254..32466
/feature="MIR repeat: matches 14. .252 of consensus"
32592..32859
/feature="AluSC repeat: matches 1. .291 of consensus"
32937..32978
/feature="L2 repeat: matches 2646. .2691 of consensus"
37001..37314
/feature="AluX repeat: matches 1. .302 of consensus"
37791..37847
/feature="L2 repeat: matches 2648. .2706 of consensus"
39138..39181
/feature="L2 repeat: matches 2664. .2707 of consensus"
39985..40256
/feature="AluX repeat: matches 1. .273 of consensus"
40837..41091
/feature="AluY repeat: matches 41. .295 of consensus"
41847..41867
/feature="MER57B repeat: matches 381. .403 of consensus"
41868..42320
/feature="L1MA2 repeat: matches 5843. .6308 of consensus"
42321..42472
/feature="MER57B repeat: matches 250. .381 of consensus"
42331..423518
/feature="multiple ESTs; match: 5' EST H94135 clone
242834; Paired with EST H94060 matching this clone; match:
3' EST H94060 clone 242834; Paired with EST H94135
matching this clone"
42343..42499
/feature="MER93 repeat: matches 216. .371 of consensus"
42574..42644
/feature="MER93 repeat: matches 18. .81 of consensus"
42651..42941
```

Query Match	46.98;	Score 146.2;	DB 9;	Length 158557;
Best Local Similarity	81.68;	Pred. No. 2.8e-22;		
Matches 169;	Conservative	0;	Mismatches 38;	Indels 0;
				Gaps 0;

RESULT	6
AB062458S2	
LOCUS	AB062458S2
DEFINITION	AB062458S2 822 bp DNA linear PR1 07-JUN-2001
VERSION	Pan troglodytes F9 gene for coagulation factor XI, exon 2 and exon 3, isolate:504.
ACCESSION	AB062460
KEYWORDS	AB062460 GI:14270093
SEGMENT	2 of 7
SOURCE	Pan troglodytes (isolate:504) female DNA.
ORGANISM	Pan troglodytes

JOURNAL Submitted (29-MAY-2001) Yoko Satta, The Graduate University for

FEATURES	Location/Qualifiers
source	1. .822

exon

ORIGIN

163 tcacagaatttggctccatgcccataagagaatttgcttcagatatattggaataa 222

db

SOURCE

ORGANISM

JOURNAL

COMMENT

```
Location/Qualifiers
1. .599
/organism="Homo sapiens"
```

```

/db.xref="taxon:9606"
/map="Xq26.3-q27.1"
prim_transcript <1..>599
/gene="F9"
/note="fix mRNA"
<1..175
/gene="F9"
/note="fix intron 1"
49..50
/gene="F9"
/citation={1}
176..339
/gene="F9"
/note="G00-119-900"
/number=2
340..527
/gene="F9"
/note="fix intron 2"
528..552
/gene="F9"
/number=3
553..>599
/gene="F9"
/note="fix intron 3"

BASE COUNT      217 a      201 t
ORIGIN            81 c     100 g
About 5.4 kb after segment 1; chromosome Xq26.3-q27.2.

```

Query Match	Similarity	45.88;	Score 142.8;	DB 9;	Length 599;
Best Local	Similarity	92.68;	Pred. No. 2.9e-21;		
Matches 150;	Conservative	0;	Mismatches 12;	Indels	Gaps 0;
Qy	148	agtaactggtgggaacacatccacagatttggctgcacatcccttaagaagaanaatggcttcacaga	207		
Db	18	AGTACTGTGGGGAACATCATCAGATTTTGGCTGCATGSCCTTAAGAAGAATTTGGCTTTCAGA	77		
Oy	208	ttatttggattcaaaaacaaagaacattctctctaagaagatgtgtaaaatttcgatggtttctt	267		
Db	78	TTATTGTGATTTAAAAACAAAGACCTTTCTTAAGAGATGTAAAAATTTTCATGATGTGTTTCTT	137		
Oy	268	ttttgciaaacataagaattaacgcgactctctttacatcatt	309		
Db	138	TTTTGCTAAACCTAAGAATTTATTTCTTTTACATTTTCAATTTT	179		
RESULT	8				
LOCUS	HUMFIXG1				
DEFINITION	Human factor IX gene, exon 1.	600 bp	DNA	linear	PRI 01-DEC-1994
ACCESSION	K02048				
VERSION	K02048.1	GI:182614			
KEYWORDS	Christmas factor; factor IX.				
SOURCE	1 of 6				
ORGANISM	Human; cDNA to liver mRNA, clones cVII, cVI, 108.1, and DB.1; 4X lymphoblastoid cell line (GM4146) DNA, clone lambda-HIX-4; genomic DNA library of Lam et al., clones lambda-HIX-1,2,3.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 600)				
JOURNAL	Anderson, D.S., Choo, K.H., Rees, D.J., Giannelis, F., Gould, K., Huddleston, J.A. and Brownlee, G.G.				
MEDLINE	The gene structure of human anti-haemophilic factor IX				
REVIEW	EMBO J. 3 (5), 1053-1060 (1984)				
AUTHORS	84236100				
COMMENT	2 (bases 138 to 139; 309 to 310; 319 to 319)				
	Anderson, D.S., Choo, K.H., Rees, D.J.G., Giannelis, F., Gould, K., Huddleston, J.A. and Brownlee, G.G.				
	Unpublished (1985)				
	[2] revision of [1].				
	The factor IX gene is about 34 kb long, and divided into 8 exons.				
	The introns range in size from .2 kb to 10 kb. [1] sequenced both				
	the cDNA for the complete mRNA and all the genomic exons, and found				

FEATURES		no evidence for polymorphism between the mRNA and the gene.	
source	Location/Qualifiers		
	1. .600		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="Xq26.3-q27.1"		
old_sequence	138..139		
	/citation=[1]		
prim_transcript	296..>600		
	/note="fix mRNA"		
old_sequence	309..310		
	/citation=[1]		
old_sequence	319		
	/citation=[1]		
	/citation=[1]		
	<325..412		
exon	/gene="F9"		
	/note="factor IX; G00-119-900"		
	/number=1		
intron	413..>600		
	/gene="F9"		
	/note="fix intron 1"		
BASE COUNT	168 a 134 c 113 g 185 t		
ORIGIN	123 bp upstream of BglI site; chromosome Xq26.3-27.2.		

Query Match	Similarity	45.6%	Score	142.2	DB	9	Length	600	
Best Local	Similarity	98.0%	Pred.	No.	3.9e-21				
Matches	144:	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
Oy	8	tatgatccctttttaaatacatcagtagtgcgtgacctttagatatagaataatctga	67						
Dd	414	TTTGTTCCCTTTTAAAAATACATGAGATGCTGGCCTTTAGATATTGAATAATATCTGA	473						
Oy	68	tgcctctcttcactcaatttttgatttacatgatttgacaagaataatgaagctaac	127						
Dd	474	TGCCTCTCTTCCTCACTAAATTTTGATTCATGATTGTGACAGCAAAATTGAAGAAGCTAAC	533						
Oy	128	aggcacgcacgagcttggtlaagtactg	154						
Dd	534	AGCGACGACGACGAGTTGGTAAGTACTG	560						
RESULT	9								
LOCUS	AB062459S2								
DEFINITION	Pan troglodytes F9 gene for coagulation factor XI, exon 2 and exon								
ACCESSION	AB062461								
VERSION	AB062461.1								
KEYWORDS	GI:14270117								
SEGMENT									
SOURCE									
ORGANISM	Pan troglodytes (isolate:505) male DNA.								
REFERENCE									
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.								
JOURNAL	1 (sites)								
REFERENCE	Comparison of DNA and protein polymorphisms between humans and								
AUTHORS	Chimpanzees								
TITLE	Genes Genet. Syst. (2001) In press								
JOURNAL	2 (bases 1 to 822)								
REFERENCE	Saitta, Y.								
AUTHORS	Direct Submission								
TITLE	Submitted (29-MAY-2001) Yokoi Saitta, The Graduate University for								
JOURNAL	Advanced Studies, Department of Biosystems Science; 1560-35								
FEATURES	Kamiyamauchi, Hayama, Kanagawa 240-0193, Japan								
SOURCE	(E-mail:saitta@mailsv.soken.ac.jp, Tel:81-468-58-1574,								
	Fax:81-468-58-1544)								
	Location/Qualifiers								
	1..822								
	/organism="Pan troglodytes"								
	/isolate="505"								
	/db_xref="taxon:9598"								

exon	/sex="male"		/note="CDS is reported in Acc#AB062471"	
	352..515			
	/gene="Fg"			
	/number=2			
	/product="coagulation factor XI"			
exon	704..728			
	/gene="Fg"			
	/number=3			
	/product="coagulation factor XI"			
BASE COUNT	293 a 110 c 130 g 287 t		2 others	
ORIGIN				
Query Match	45.5%	Score 142;	DB 9;	Length 822;
Best Local Similarity	80.2%	Pred. No. 4.1e-21;		
Matches 166;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;
OY	103	tgacagcaatattgtaaggtcctaacaagccagacgcaggttgtaagtctgggaaca	162	
Db	149	TAAAGTAAATTTGAATTTTAACTCTTAATCTCATGTGATACAGAGACTGTGGAAACA	208	
OY	163	tcaacgattttggtcccatgccttaagaagaattggtcttcagattatttgattaaa	222	
Db	209	TCACAGATTTTGGATCCAGCCCTAAAGAGAATTTGGCTTCAGATTATTGGATTAAA	268	
OY	223	acaaagacttctctaagagatgtaaatcttcagtgttcttcttttgcctaaactaa	282	
Db	269	ACAAAGACTTCTTTTAAGAGATGTAATAATTTTCATGATGTTTCTTTTNGCTAAACATRA	328	
OY	283	agaattaacgcgtatctttacattt	309	
Db	329	AGAATTATTTCTTTTACATTTTCAGTTT	355	
RESULT 10				
AB062458S1				
LOCUS	AB062458S1	347 bp	DNA	linear
DEFINITION	Pan troglodytes F9 gene for coagulation factor XI, exon 1,			
	isolate:504.			
ACCESSION	AB062458			
VERSION	AB062458.1	GI:14270092		
KEYWORDS				
SEGMENT				
SOURCE				
ORGANISM				
	i of 7			
	Pan troglodytes (isolate:504) female DNA.			
	Pan troglodytes			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.			
REFERENCE				
AUTHORS	Satia.Y.			
TITLE	Comparison of DNA and protein polymorphisms between humans and			
	Chimpanzees			
JOURNAL	Genes Genet. Syst. (2001) In press			
REFERENCE	2 (bases 1 to 347)			
AUTHORS	Satia.Y.			
TITLE	Direct Submission			
JOURNAL	submitted (29-MAY-2001) Yoko Satia, The Graduate University for			
	Advanced Studies, Department of Biosystems Science; 1560-35			
	Kamlyamaguchi, Hayama, Kanagawa 240-0193, Japan			
	(E-mail:satia@mailsv.soken.ac.jp, Tel:81-468-58-1574,			
	Fax:81-468-58-1544)			
FEATURES				
source	Location/Qualifiers			
	1..347			
	/organism="Pan troglodytes"			
	/isolate="504"			
	/db_xref="taxon:9598"			
	/sex="female"			
	/note="CDS is reported in Acc#AB062470"			
5' UTR	108..136			
	/gene="Fg"			
exon	108..224			
	/gene="Fg"			
	/number=1			

	BASE COUNT	104	a	71	c	64	g	108	t	/product="coagulation factor XI"	
	ORIGIN										
	Query Match	37.6%: Score 117.2; DB 9; Length 347;									
	Best Local Similarity	97.5%; Pred. No. 1e-15;									
	Matches 119;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;		
OY	8	tatgcacccctttttaaatatcatgtgatgctgcctttagatatagaataatcga	67								
Dd	226	TTTGTTTCCTTTTAAATAACATGAGATGCTGCCTTTAGATATGAATAATCTCA	285								
OY	68	tctgtcttcctcaacaaatttgttatcatatttgacaagaataatgaagctaac	127								
Dd	286	TGCTGTCTTCCTCACTAATATTGGATTAATATTTGAACAGCAATATTGAAGAAGCTAAC	345								
OY	128	ag	129								
Dd	346	AG	347								
	RESULT 11										
	AB062459S1	347	Dp	DNA	linear	PRI 31-MAY-2001					
	LOCUS	Pan troglodytes F9 gene for coagulation factor XI, exon 1,									
	DEFINITION	isolate:505.									
	ACCESSION	AB062459									
	VERSION	AB062459.1 GI:14270116									
	KEYWORDS										
	SEGMENT										
	SOURCE										
	ORGANISM	1 of 7 Pan troglodytes (isolate:505) male DNA. Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.									
	REFERENCE										
	AUTHORS	Satla,Y. Comparison of DNA and protein polymorphisms between humans and chimpanzees									
	TITLE	Genes genet. Syst. (2001) In press									
JOURNAL	REFERENCE	2 (bases 1 to 347)									
AUTHORS	TITLE	Direct Submission									
JOURNAL	TITLE	Submitted (29-MAY-2001) Yoko Satla, The Graduate University for Advanced Studies, Department of Biosystems Science; 1560-35 Kamlayamaguchi, Hayama, Kanagawa 240-0193, Japan (E-mail:satla@mailsy.soken.ac.jp, Tel:81-468-58-1574, Fax:81-468-58-1544)									
FEATURES	Location/Qualifiers										
SOURCE	1..347 /organism="Pan troglodytes" /isolate="505" /db_xref="taxon:9598" /sex="male" /note="CDS is reported in Acc#:AB062471" 5' UTR /gene="F9" 108..136 exon 108..224 /gene="F9" /number=1 /product="coagulation factor XI" BASE COUNT 104 a 71 c 64 g 108 t ORIGIN										
QY	8	tatgcacccctttttaaatatcatgtgatgctgcctttagatatagaataatcga	67								
Dd	226	TTTGTTTCCTTTTAAATAACATGAGATGCTGCCTTTAGATATGAATAATCTCA	285								
OY	68	tctgtcttcctcaacaaatttgttatcatatttgacaagaataatgaagctaac	127								
Dd	286	TGCTGTCTTCCTCACTAATATTGGATTAATATTTGAACAGCAATATTGAAGAAGCTAAC	345								
OY	128	ag	129								
Dd	346	AG	347								

OY	68	tgcgtcttcttcacaaatttgatgacattggcagcaaatctgaagaagtctaac	127
Db	286	TGCTGTCTTCTTCACAAATTGGATTAACATGATTTGCAGCAGAATATTGAAGAGTCITAAC	345
OY	128	ag 129	
Db	346	AG 347	
RESULT	12		
AX347024			
LOCUS	AX347024	10329 bp	DNA
DEFINITION	Sequence 2095 from Patent WO0200928.		linear PAT 01-FEB-2002
ACCESSION	AX347024		
VERSION	AX347024.1	GI:18494910	
KEYWORDS	.		
SOURCE	synthetic construct.		
ORGANISM	artificial sequence.		
REFERENCE	1 (sites)		
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.		
TITLE	Diagnosis of diseases associated with the immune system		
JOURNAL	Patent: WO 0200928-A 2095 03-JAN-2002;		
Pigenomics AG (DE)			
FEATURES			
source	Location/Qualifiers		
	1..10329		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
BASE COUNT	3676 a	56 c	2107 g 4490 t
ORIGIN			
Query Match	36.9%; Score 115; DB 6; Length 10329;		
Best Local Similarity	70.3%; Pred. No. 2.1e-15;		
Matches 154; Conservative	0; Mismatches 65; Indels 0; Gaps 0;		
OY	86	atttgattcacatgatggcagcaaatctgaagagctcaacgccagccagcgacttgg	145
Db	10109	ATTGGAATTTTTTGATTTAAAGCAATAATTGAATTTTAATTTTAAATTTTATGCTAT	10168
OY	146	taagtactgtggagacatcacagatttggctccatgccttaagaagaattggctcca	205
Db	10169	ATACTATTGTGGCAATTTTATAGATTTTGGTTTATGCTTTTAAAGCAAATTCGTTTAA	10228
OY	206	gatatattggattaacaacaagaccttctaagaagatgtaaaaatttcatagatgtttc	265
Db	10229	GATTATTGTGATTTAAAATTAAGATTTTTTTTAAGAGATGTAATAATTTTATGATGTTTTT	10288
OY	266	tttttgcctaaaaaagaattaagcgcatcttta	304
Db	10289	TTTTTTGTAAAAATTAAGATTAATTTTATATATTTTA	10327
RESULT	13		
AX347025/c			
LOCUS	AX347025	10329 bp	DNA
DEFINITION	Sequence 2096 from Patent WO0200928.		linear PAT 01-FEB-2002
ACCESSION	AX347025		
VERSION	AX347025.1	GI:18494911	
KEYWORDS	.		
SOURCE	synthetic construct.		
ORGANISM	artificial sequence.		
REFERENCE	1 (sites)		
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.		
TITLE	Diagnosis of diseases associated with the immune system		
JOURNAL	Patent: WO 0200928-A 2096 03-JAN-2002;		
Pigenomics AG (DE)			
FEATURES			
source	Location/Qualifiers		
	1..10329		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		

BASE COUNT	2681 a	56 c	1865 g	5727 t
ORIGIN				
Query Match	33.2%; Score 103.6; DB 6; Length 10329;			
Best Local Similarity	70.2%; Pred. No. 6e-13;			
Matches 139; Conservative	0; Mismatches 59; Indels 0; Gaps 0;			
OY	92	attacatgatttacagcaaatctgaagagctcaacaccagcagcgattgtagta	151	
Db	215	ATTCTTTAACAATAAATAAATTTTAATTCCTTAATCCATCATATATACATA	156	
OY	152	ctctgggaacatacagatttggctccatgccttaagaagaattggcttcagattat	211	
Db	155	CTATTAACATCACCAATTTTACTTCATCCCTTAATAAATAATTAATTCCAATAT	96	
OY	212	ttgatttaaaacaagaaccttctctaagatgtaaatcttcatagatgtttctttt	271	
Db	95	TTAAATTTAAACAAACCTTCTTAAATAATTAATTTTCATATATTTCTTTT	36	
OY	272	gctaaactaagaatta	289	
Db	35	ACTAAACTTAATAATTA	18	
RESULT	14			
AL671984				
LOCUS	AL671984	221406 bp	DNA	
DEFINITION	Mus musculus chromosome X clone RP23-419J18, *** SEQUENCING IN PROGRESS ***, in unordered pieces.			
ACCESSION	AL671984			
VERSION	AL671984.1	GI:18643903		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	McLay,K.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk			
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquerry@sanger.ac.uk ----- Project Information Center project name: BM419J18 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 215188 bases at least Q40 Consensus quality: 217184 bases at least Q30 Consensus quality: 218499 bases at least Q20 Insert size: 219606; sum-of-contigs Insert size: 216941; 3.5% error; agarose-fp Quality coverage: 6.37x in Q20 bases; sum-of-contigs Quality coverage: 6.61x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. Location/Qualifiers 1..221406 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="X" /clone="RP23-419J18" /clone_1fb="RPCL-23"			

```
misc_feature 1..9656 /note="assembly_fragment:00681  
fragment_chain:1"  
misc_feature 9757..15436 /note="assembly_fragment:03363  
fragment_chain:1"  
misc_feature 15537..19499 /note="assembly_fragment:01895  
fragment_chain:1"  
misc_feature 19600..34940 /note="assembly_fragment:02973  
fragment_chain:1"  
misc_feature 35041..45481 /note="assembly_fragment:00981  
fragment_chain:1"  
misc_feature 45582..58736 /note="assembly_fragment:02515  
fragment_chain:1"  
misc_feature 58837..79679 /note="assembly_fragment:03370  
fragment_chain:1"  
misc_feature 79780..81913 /note="assembly_fragment:03258  
fragment_chain:1"  
misc_feature 82014..12076 /note="assembly_fragment:01480  
fragment_chain:2"  
misc_feature 122177..125261 /note="assembly_fragment:01088  
fragment_chain:2"  
misc_feature 125362..157061 /note="assembly_fragment:00632  
fragment_chain:2"  
misc_feature 157162..172673 /note="assembly_fragment:01181  
fragment_chain:2"  
misc_feature 172774..186159 /note="assembly_fragment:01899  
fragment_chain:2"  
misc_feature 186260..194735 /note="assembly_fragment:02258  
fragment_chain:3"  
misc_feature 194836..199785 /note="assembly_fragment:02783  
fragment_chain:3"  
misc_feature 199886..201998 /note="assembly_fragment:01513  
fragment_chain:3"  
misc_feature 202099..207922 /note="assembly_fragment:00844.0"  
misc_feature 208023..219280 /note="assembly_fragment:00844.1"  
misc_feature 219381..221406 /note="assembly_fragment:00904"  
BASE COUNT 66086 a 43516 c 43229 g 66769 t 1806 others  
ORIGIN
```

AX345014	RESULT 15				
LOCUS	AX345014	9927 bp	DNA	linear	PAT 01-FEB-2002
DEFINITION	Sequence 85 from Patent WO0200928.				
ACCESSION	AX345014				
VERSION	AX345014.1	GI:18492900			
KEYWORDS	.				
SOURCE	synthetic construct.				
ORGANISM	synthetic construct				
	artificial sequence.				
REFERENCE	1 (sites)				
AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.				
TITLE	Diagnosis of diseases associated with the immune system				
JOURNAL	Patent: WO 0200928-A 85 03-JAN-2002;				
	Epigenomics AG (DE)				
FEATURES	Location/Qualifiers				
SOURCE	1..9927				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="chemically treated genomic DNA (Homo sapiens)"				
BASE COUNT	3017 a	50 c	1946 g	4914 t	
ORIGIN					

Query Match	16.6%	Score 51.8	DB 6	Length 9927
Best Local Similarity	49.4%	Pred. No. 0.091		
Matches 134	Conservative 0	Mismatches 137	Indels 0	Gaps 0
QY	2	tttggttatgcacccctttttaaaatcacatgagtaagctgcctctttatagatagaat	61	
DB	5467	TTTTAATTTTTAATTTTTTTTTTATATATATATATTTTGTGATTTATTTATTTATTT	5526	
OY	62	atcigatcgtctctcttcacctaatttggatcacatbatttgaacgaacatctcaag	121	
DB	5527	ATTGTTTATGCTTAATTTTTTTTATTTTTTTTAAATTATTTATGATGATTTGAACTT	5586	
OY	122	tctaacagccagcacgcaggttbgtaagctactblyggaaacatcaagaatttggctccat	181	
DB	5587	TTTTTGTATAGAAATAAATTTTTGGACATATTTTGTTTGGTTTATTTATTTTAA	5646	
OY	182	gocccaaagagaatcttgctcttcagatatttggatgaataaacaagaactcttaag	241	
DB	5647	ATGAGGAATGTAAATAGGCTTGGAATTTTATAGAGTACTTATTTATTTATATGAT	5706	
OY	242	atgtaaaatttccatgatgtcttctttttg	272	
DB	5707	TTTTTAAATTTTTTATGTTGATATTTTATG	5737	

Search completed: September 9, 2002, 21:17:53
Job time: 17751 sec

Query Match	Similarity	21.6%	Score 67.4	DB 2	Length 221406
Best Local	Similarity	69.3%	Pred. No. 2.7e-05		
Matches	106	Conservative	0	Mismatches 46	Indels 1
				Gaps 1	
QY	159	aacatcacagatttgcgtccatgccttaagaagaatgagcttcagatttgatt	218		
Db	122765	AAATCATATGTTTGGCTCTATGCCCAAAAGAAATTAAGTATGGAATTTACATGAAC	122824		
QY	219	aaaaacaaagacttccttaagagatgta-aaattcacagatgtttcttcttgcgttaa	277		
Db	122825	CAAAACAAACCTTTCTTTAAAAATATTATGTTTTCGAGGTTTTTTTTTTCCTAAT	122884		
QY	278	actaaagaattaaagcgctattcttttacattc	310		
Db	122885	ACTAAAGAACATTAACCTTTAAATTTTCAGTTTTC	122917		

•
•
•
•